John von Neumann Institute for Computing

Nonadiabatic Car-Parrinello Molecular Dynamics Study of the Tautomerism of DNA Bases

H. Langer, N. L. Doltsinis, D. Marx

published in

NIC Symposium 2004, Proceedings,
Dietrich Wolf, Gernot Münster, Manfred Kremer (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **20**, ISBN 3-00-012372-5, pp. 81-89, 2003.

http://www.fz-juelich.de/nic-series/volume20

Nonadiabatic Car-Parrinello Molecular Dynamics Study of the Tautomerism of DNA Bases

H. Langer, N. L. Doltsinis, and D. Marx

Lehrstuhl für Theoretische Chemie, Ruhr–Universität Bochum, 44870 Bochum, Germany
E-mail: holger.langer@theochem.ruhr-uni-bochum.de, nikos.doltsinis@rub.de
dominik.marx@theochem.ruhr-uni-bochum.de

Photophysical properties of the DNA base guanine in the lowest singlet excited electronic state, S_1, both in the gas phase and in aqueous solution have been investigated using the density functional restricted open-shell Kohn-Sham method. The biologically relevant N9H-keto tautomer exhibits significantly larger structural changes than other energetically close-lying tautomers. This excited state distortion is site-specifically enhanced by methylation at position N9 offering a plausible explanation for the absence of an optical absorption signal for N9Methyl guanine. First principles molecular dynamics simulations have shown that guanine does not spontaneously tautomerize at room temperature neither in the gas phase nor in aqueous solution. Tautomerisation mechanisms have therefore been studied using coordination constrained dynamics.

1 Introduction

The purines guanine (G) and adenine (A), on one hand, and the pyrimidines cytosine (C) and thymine (T), on the other, are complementary bases forming the hydrogen-bonded G-C and A-T Watson-Crick base pairs in DNA[1–4] whose sequence stores the genetic code. Although, in the gas phase, all nucleic acid bases can exist in a variety of energetically close-lying tautomeric forms, only one tautomer of each species is present in the canonical base pairs of healthy DNA[1–4]. It is one of the most fundamental goals in biochemistry to shed light on the underlying natural selection process. In the case of G, for example, the ground state energies of the four most stable tautomers have been calculated to lie within a narrow range of 7 kJ/mol[5,6], but only the N9H-*keto* form (see Fig. 1 for structure and nomenclature of this and other G tautomers discussed in the following) is biologically relevant.

Studying *isolated* nucleic acid bases as opposed to in their biological environment offers several crucial advantages. Firstly, gas phase spectra are sparser and thus much easier to interpret, rid of the complex interactions with the backbone and the solvent. Secondly, meaningful *ab initio* electronic structure calculations on large biological systems are not yet feasible because the computational cost is prohibitive. Research on isolated DNA bases, on the other hand, can already benefit from fruitful interplay between experiment and theory. Gas phase resonant two-photon ionization ultraviolet (UV) spectroscopy[6–15] of electronically excited nucleobases has become possible by laser desorption and supersonic jet expansion techniques avoiding extensive decomposition during vaporization. Fluorescence lifetimes have been measured by means of femtosecond spectroscopy[16–19].

Knowledge of the photophysical properties of nucleic acid bases is the key to understanding radiation induced genetic damage[19–22]. It is generally believed that the evolutionary process has selected those nucleobase tautomers as DNA building blocks that have the shortest excited state lifetimes in order to prevent any damaging photochemical reactions[10,15,19].

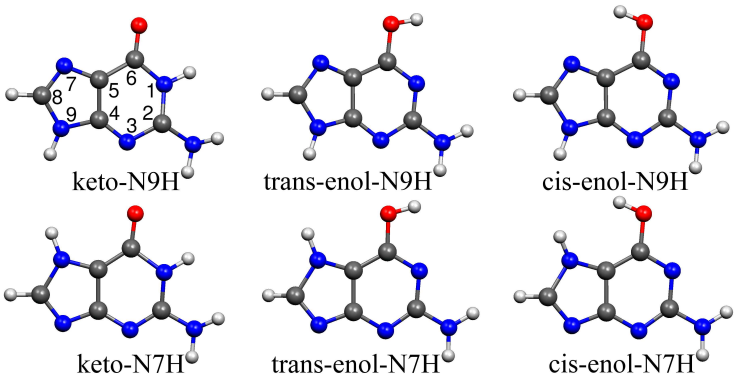

Figure 1. Structure and nomenclature of the six most stable G tautomers (colour code: red=oxygen, blue=nitrogen, grey=carbon, white=hydrogen). The atomic numbering scheme is illustrated for the N9H-*keto* tautomer.

It is the aim of this project to calculate structural, energetic, and spectroscopic quantities of the six lowest energy G tautomers (Fig. 1) in the S_1 excited state using the new restricted open-shell Kohn-Sham (ROKS)[23] density functional approach. The dynamics of G tautomerisation in the S_1 state shall be studied both in the gas phase and in aqueous solution by means of the Car-Parrinello molecular dynamics (CP-MD) method[24,25] including the state-of-the-art nonadiabatic CP-MD[26–28] and coordination constraint techniques[29–31].

2 Gas Phase Tautomerism

Since this part of the project has already been published[32,33], we confine ourselves to the essentials and refer the reader to Ref. 32 for more details.

Using the BLYP functional we have found the N7H-*keto* G to be the most stable tautomer both in the ground state and in the first excited $\pi \rightarrow \pi^*$ state, slightly lower in energy than the N9H-*keto* tautomer (see Table 1 for a summary of various energetic quantities computed in this work). We observe substantial geometrical distortions in the S_1 state compared to the ground state, in particular for the N9H-*keto* tautomer whose six-membered aromatic ring is heavily nonplanar.

Our theoretical adiabatic S_1 excitation energies (Table 1) can be compared to experimental 0–0 transition energies providing hints as to the spectral positions of the individual G tautomers. In combination with our ROKS S_1 vibrational spectra, the present results facilitate the assignment of experimental IR-UV and REMPI spectra of jet-cooled G. In addition, this work has demonstrated that excited state vibrational frequencies can be obtained fairly reliably using the ROKS method. In particular, unlike the more conventional CIS and CASSCF methods, ROKS does not require any rescaling of vibrational frequencies. Velocity autocorrelation functions obtained from adiabatic excited state CP-MD simulations demonstrate that anharmonic effects only play a minor role.

Besides the characterisation of the individual G tautomers an important goal of this project has been the investigation of the tautomerisation mechanism. We have, therefore, carried out CP-MD simulations of the N9H-*enol* tautomer in the gas phase at various tem-

tautomer	E^0	E^1	$\Delta E^{\rm vert}$	$\Delta E^{\rm ad}$
N7H-*keto*	0.0	0.0	3.35	2.97
N9H-*keto*	3.3	16.3	3.70	3.11
trans-N9H-*enol*	12.8	22.5	3.52	3.07
cis-N9H-*enol*	14.4	20.7	3.50	3.04
trans-N7H-*enol*	23.0	35.0	3.48	3.10
cis-N7H-*enol*	52.2	63.3	3.47	3.09

Table 1. Relative energies in kJ/mol of the six most stable G tautomers in the ground state, E^0, and in the S_1 excited state, E^1, calculated using the BLYP functional in a plane-wave basis truncated at 100 Ry. ROKS vertical and adiabatic excitation energies (given in eV) are labeled $\Delta E^{\rm vert}$ and $\Delta E^{\rm ad}$, respectively.

peratures. Our conclusion is that spontaneous tautomerisation involving proton transfer is not observable in the time window of a few picoseconds permitted by *ab initio* MD even at increased temperature as high as 1000 K. However, we do observe frequent *cis–trans* isomerisation events suggesting that the two *enol* isomers are indistinguishable experimentally.

3 Methylated Guanine Tautomers

Replacing a hydrogen atom in G by a methyl group has been shown experimentally to have only minor effects on the S_1 optical absorption spectra for most tautomers. A notable exception is the 9Me-*keto* tautomer which has so far proven impossible to detect. Our ROKS calculations suggest that the excited state global minimum structure of 9Me-*keto* G is heavily distorted compared to the ground state and therefore the probability for optical absorption should be low. In contrast to the other tautomers investigated, e.g. 7Me-*keto* G, the S_1 geometry of 9Me-*keto* G is vastly different from the respective unmethylated species. Table 2 summarizes structural and energetic parameters of the 7Me-*keto* and 9Me-*keto* tautomers as well as their unmethylated counterparts. A measure for the degree of structural distortion in the S_1 state is the root mean squared distance (rmsd) between ground and excited state geometries. The rmsd values listed in Table 2 illustrate that geometric changes following photoexcitation are selectively enhanced by 9-methylation. In order to eliminate any contributions from rotation of the methyl group, we have recomputed the rmsd values of both methylated *keto* tautomers after replacing the methyl group by a hy-

tautomer	E^0	E^1	$\Delta E^{\rm vert}$	$\Delta E^{\rm ad}$	$\Delta E^{\rm fl}$	rmsd
N7H-*keto*	0.0	0.0	3.35	2.97	2.56	0.057
N9H-*keto*	3.3	16.3	3.70	3.11	1.97	0.109
N7Me-*keto*	0.0	0.0	3.28	2.89	2.47	0.075 (0.053)
N9Me-*keto*	2.7	17.3	3.67	3.05	1.72	0.230 (0.237)

Table 2. Relative energies in kJ/mol of *keto* G tautomers in S_0 and S_1, E^0 and E^1. The comparison is among unmethylated and methylated species, respectively. Vertical and adiabatic excitation as well as fluorescence energies, $\Delta E^{\rm vert}$, $\Delta E^{\rm ad}$, and $\Delta E^{\rm fl}$ are given in eV. Rmsd's are in Å (values in parentheses for substitution of CH_3 with H, see text).

drogen atom (see values in parentheses in Table 2). This way, the rmsd results for N7H and N7Me G are almost identical, indicating that methylation has little impact on the general molecular structure of the N7-substituted *keto* form. The situation is very different for N9Me-*keto* G; in this case the CH_3 group is seen to make only a minor contribution to the rmsd value. Here, in-plane distortions of the six-membered ring as well as out-of-plane distortions of the amino group, i.e. skeletal rearrangements, are chiefly responsible for the large overall deviation. We should stress that the S_1 global minimum geometries of N9Me and N9H are distinctly different, the six-membered ring of N9H exhibiting large out-of-plane distortions whereas N9Me is essentially flat except for the heavy out-of-plane bending of the amino group absent in N9H. Table 2 further shows that for both methylated and unmethylated G the N9-substituted *keto* tautomers exhibit significantly larger structural rearrangement compared to the N7-substituted tautomers. A graphical representation of the different molecular structures and a schematic view of the potential landscapes can be found in Fig. 3. Going beyond the purely *static* analysis we have carried out excited state *ab initio* MD simulations of methylated G in order to study the *dynamics* of geometric relaxation after photoexcitation. The dynamical treatment has indeed been crucial in revealing the involvement of an additional, local, S_1 minimum in the structural rearrangement of N9Me G. During the first few femtoseconds both tautomers typically behave in a similar way (see Fig. 2, phase I) showing mainly in-plane bond relaxation. The N9Me tautomer, however, enters a second phase (see Fig. 2, phase II) after roughly 30 fs when

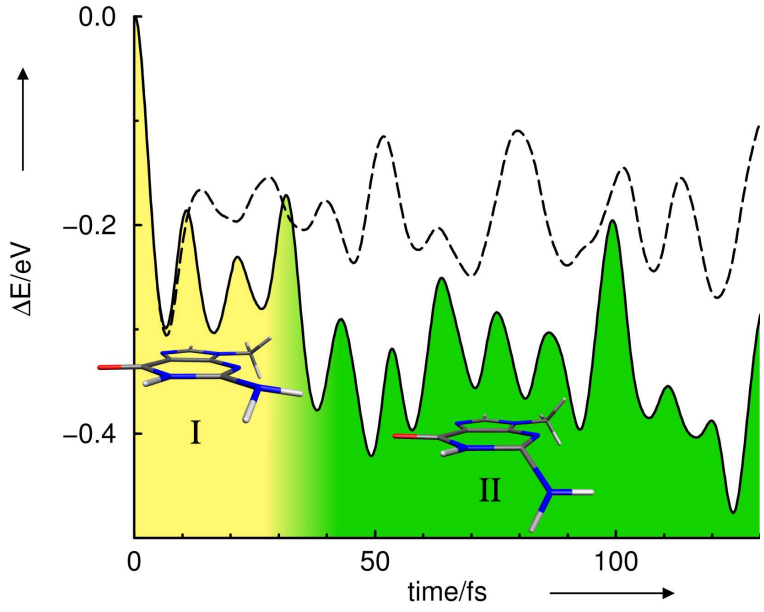

Figure 2. Typical time evolution of S_1 energy relative to the vertical excitation energy, ΔE, after vertical excitation of N9Me-*keto* (solid line) and N7Me-*keto* G (dashed line). The structural relaxation of N9Me-*keto* proceeds through two potential minima (optimized geometries are superimposed); Phase I (local minimum): predominantly in-plane relaxation, Phase II (global minimum): out-of-plane motion of the amino group.

out-of-plane ring vibrations set in and the amino group starts to bend strongly. The energy then continues to have a downward tendency over several hundred femtoseconds until the system reaches the global minimum. In a further series of MD runs at fixed temperatures between 10 and 50 K, we have been able to show that during phase I N9Me travels through a local S_1 minimum, where it can be trapped at temperatures lower than 50 K. The corresponding optimized local minimum structure lies energetically 0.15 eV above the global S_1 minimum; it is largely planar except for the C_8H_8 bond which is bent out-of-plane. In a realistic scenario, however, the system gathers too much momentum during the initial ballistic phase to remain in the local minimum and thus ends up in the geometrically distant global S_1 minimum. As a consequence, we expect the overlap between the S_0 and S_1 nuclear wavefunctions and therefore the absorption probability to be small. To clarify both the main methylation effects and their tautomer selectivity Figure 3 offers a schematic view of the potential shapes of all four tautomers together with the relevant minimum geometries.

In summary, we have performed ROKS density functional calculations to characterize the first excited singlet state of N7Me and N9Me G, the latter being a minimalistic mimetic of G linked to the DNA backbone. Adding "merely a methyl group" to G results

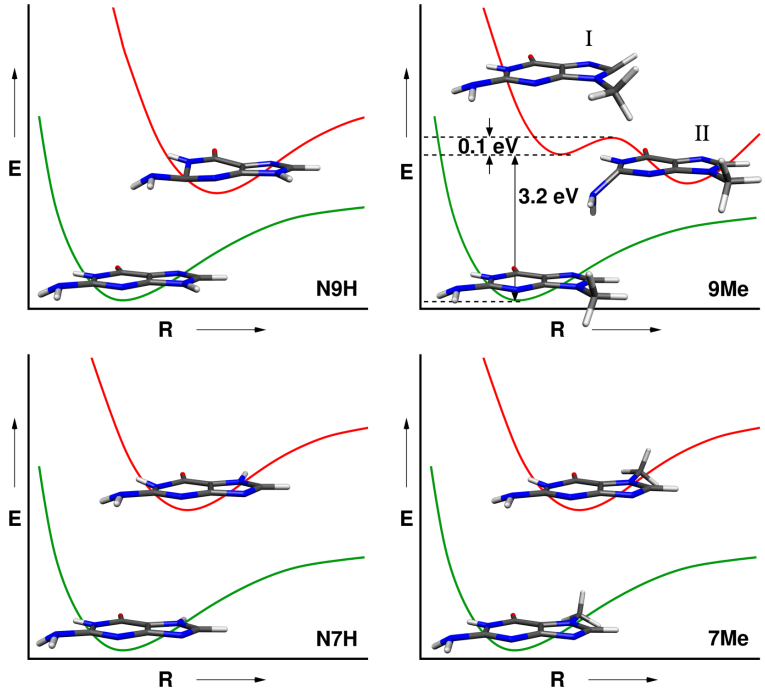

Figure 3. Schematic view of the S_0 and S_1 potential landscapes of N9H, N7H, N9Me, and N7Me G along the reaction coordinate, R, together with dynamically important minimum structures. The excited state minima I and II of N9Me G are explored during phases I and II of the MD simulation, respectively (see Figure 2). Methylation alters dramatically the S_1 potential topology in the case of N9-substituted G, whereas N7-substituted G remains largely unaffected.

in vastly different structural relaxation dynamics after vertical S_1 excitation only in the case of N9-substitution, whereas N7H G remains largely unaffected. In contrast to all other tautomers investigated, photoexcitation initially takes N9Me G to a *local* S_1 minimum. Since the latter is thermally unstable, the system then decays to the global minimum, which is geometrically far from the S_0 structure giving rise to very poor vertical excitation efficiency. We therefore expect optical absorption in this spectral region to be comparatively weak. This is corroborated by the fact that only the N7Me tautomer has been observed experimentally[9]. Methylation at N9 position clearly enhances the resistance of *keto* G with respect to photochemistry. In addition to short excited state lifetimes, the "forbidden excitation" mechanism proposed here may contribute to protecting DNA from UV damage. Apart from structural relaxation in the excited state, another important issue is that of radiationless decay of the S_1 state, which we shall address by applying our nonadiabatic CP-MD technique[26–28].

4 Tautomerisation Dynamics in Aqueous Solution

Tautomerisation in aqueous solution is expected to be enhanced by the existence of additional solvent-assisted proton transfer pathways. We have performed CP-MD calculations of *cis*-N7H-*enol* G embedded in a periodically repeated unit cell with 60 H_2O molecules both in the ground state and in the S_1 state (see Fig. 4 for a graphical representation of the simulation cell). On a time scale of roughly 4 ps, no proton transfer has occurred. Analysis

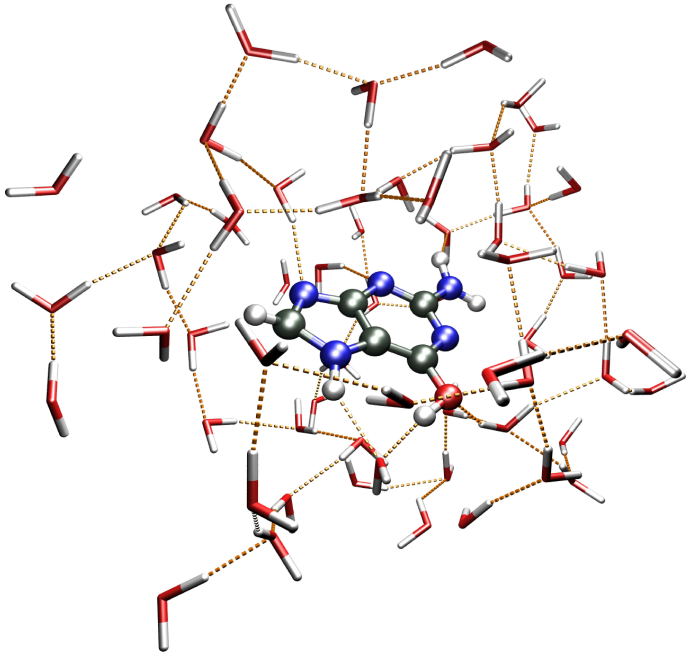

Figure 4. Snapshot of Car-Parrinello MD simulation unit cell containing *cis*-N7H-guanine solvated by 60 water molecules.

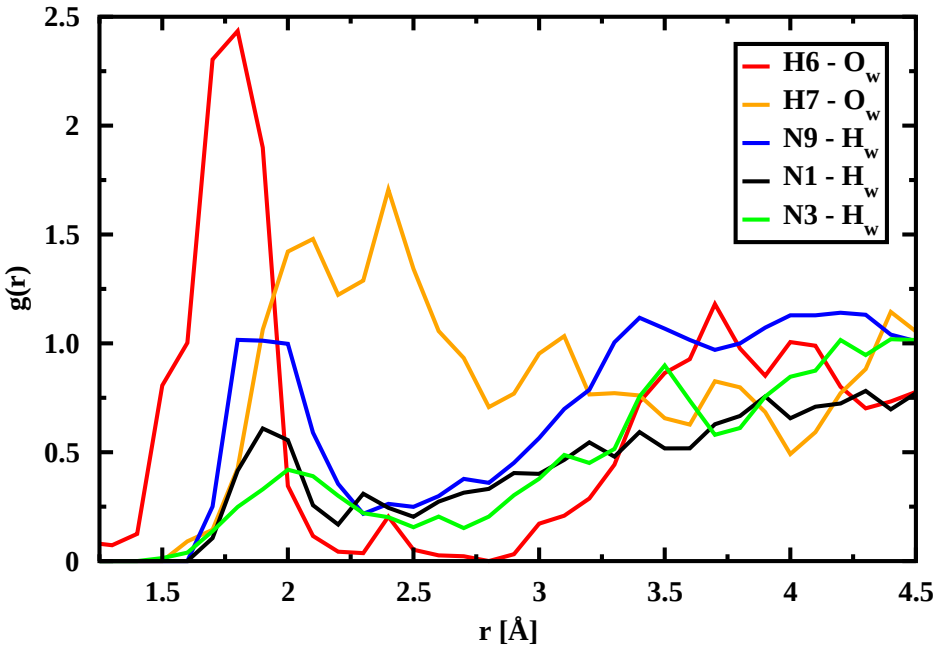

Figure 5. Radial distribution functions for the most acidic and the most basic G sites, respectively. Out of the most likely proton donors, the hydrogen bonds between hydrogen H6 of the *enol* OH group and oxygen atoms, O_w, of the solvent water molecules are seen to be shortest. Nitrogen N9, on the other hand, is clearly the most probable proton acceptor as it forms the shortest hydrogen bonds with water hydrogens, H_w.

of the excited state radial distribution functions for solute–solvent hydrogen bonds at various sites of the G molecule (see Fig. 5) reveals that nitrogen N9 is the most likely candidate for protonation, closely followed by nitrogen N1. By far the most acidic site of the G molecule seems to be the OH group followed by the N7H group. However, our calculations so far indicate that *cis*-N7H-*enol* G in aqueous solution is stable at least on a picosecond time scale. The proton donor sites must therefore be only weakly acidic, whereas the proton acceptor sites are only weakly basic.

Since no spontaneous tautomerisation can be observed during the time scale of the ab initio MD simulation, we are now studying possible proton transfer mechanisms by means of a coordination constraint technique[29,31]. Motivated by the observations during the unconstrained simulation, our first objective is to break the *enol* OH bond, which is apparently the most likely scenario. For this purpose, we incrementally reduce the coordination number of the oxygen atom from a value of approximately unity (corresponding to the unconstrained equilibrium) to zero (corresponding to complete deprotonation). Figure 6 illustrates how a hydronium ion, H_3O^+, is formed as the coordination constraint breaks the OH bond. Further reduction of the coordination number will then lead to the onset of a Grotthus-type diffusion of the proton through the water solvent. Eventually the proton is expected to recombine with the G molecule. Thus, the simulation should reveal which of the *keto* G tautomers is energetically favourable in aqueous solution.

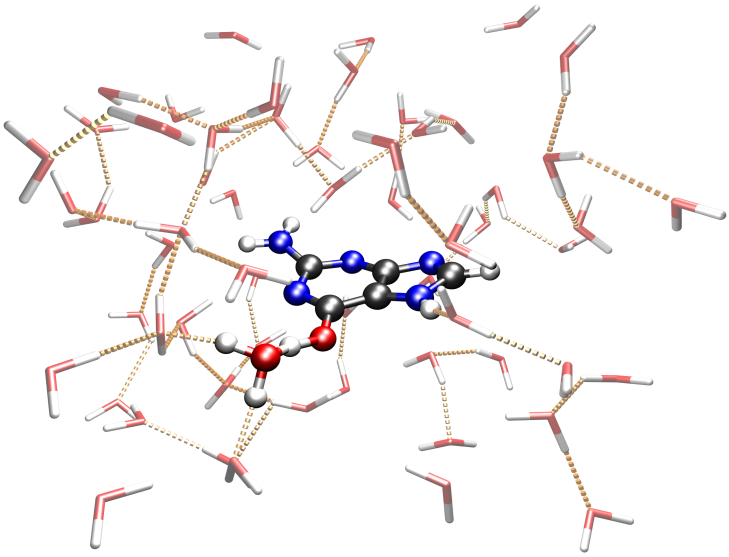

Figure 6. Snapshot of coordination constrained *cis*-N7H-guanine solvated by 60 water molecules. As the coordination number of the oxygen atom is reduced, the OH bond is broken and a hydronium ion, H_3O^+, is formed.

5 Acknowledgment

The John von Neumann Institute for Computing, Forschungszentrum Jülich is gratefully acknowledged for providing computational resources.

References

1. E. Chargaff. *Experientia*, 6:201, 1950.
2. E. Chargaff. *Experientia*, 26:810, 1970.
3. E. Chargaff. *Science*, 172:637, 1971.
4. J. D. Watson and F. H. C. Crick. *Nature*, 171:737, 1953.
5. T. K. Ha, H. J. Keller, R. Gunde, and H. H. Gunthard. *J. Phys. Chem. A*, 103:6612, 1999.
6. E. Nir, C. Janzen, P. Imhof, K. Kleinermanns, and M. S. de Vries. *J. Chem. Phys.*, 115:4604, 2001.
7. E. Nir, L. Grace, B. Brauer, and M. S. de Vries. *J. Am. Chem. Soc.*, 121:4896, 1999.
8. F. Piuzzi, M. Mons, I. Dimicoli, B. Tardivel, and Q. Zhao. *Chem. Phys.*, 270:205, 2001.
9. M. Mons, I. Dimicoli, F. Piuzzi, B. Tardivel, and M. Elhanine. *J. Phys. Chem. A*, 106:5088, 2002.
10. E. Nir, K. Kleinermanns, and M. S. de Vries. *Nature*, 408:949, 2000.
11. N. J. Kim, G. Jeong, Y. S. Kim, J. Sung, and S. K. Kim. *J. Chem. Phys.*, 113:10051, 2000.

12. D. C. Lührs, J. Viallon, and I. Fischer. *Phys. Chem. Chem. Phys.*, 3:1827, 2001.
13. C. Plützer, E. Nir, M. S. de Vries, and K. Kleinermanns. *Phys. Chem. Chem. Phys.*, 3:5466, 2001.
14. E. Nir, M. Müller, L. I. Grace, and M. S. de Vries. *Chem. Phys. Lett.*, 355:59, 2002.
15. E. Nir, K. Kleinermanns, L. Grace, and M. S. de Vries. *J. Phys. Chem. A*, 105:5106, 2001.
16. A. H. Zewail. *Angew. Chem. Int. Ed.*, 39:2586–2631, 2000.
17. J. Peon and A. H. Zewail. *Chem. Phys. Lett.*, 348:255, 2001.
18. S. K. Pal, J. Peon, and A. H. Zewail. *Chem. Phys. Lett.*, 363:363, 2002.
19. J.-M. L. Percourt, J. Peon, and B. Kohler. *J. Am. Chem. Soc.*, 123:10370, 2001.
20. O. D. Schärer. *Angew. Chem.*, 115:3052–3082, 2003.
21. J. D. Watson and F. H. C. Crick. *Nature*, 171:964, 1953.
22. M. K. Cichon, S. Arnold, and T. Carell. *Angew. Chem. Int. Ed.*, 41:767–770, 2002.
23. I. Frank, J. Hutter, D. Marx, and M. Parrinello. *J. Chem. Phys.*, 108:4060, 1998.
24. D. Marx and J. Hutter. In J. Grotendorst, editor, *Modern Methods and Algorithms of Quantum Chemistry*. NIC, Jülich, 2000. for downloads see `http://www.theochem.ruhr-uni-bochum.de/go/cprev.html`.
25. `CPMD 3.4`: J. Hutter, P. Ballone, M. Bernasconi, P. Focher, E. Fois, S. Goedecker, D. Marx, M. Parrinello, and M. Tuckerman; MPI für Festkörperforschung, Stuttgart and IBM Zurich Research Laboratory.
26. N. L. Doltsinis and D. Marx. *Phys. Rev. Lett.*, 88:166402, 2002.
27. N. L. Doltsinis. In J. Grotendorst, D. Marx, and A. Muramatsu, editors, *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*. NIC, FZ Jülich, 2002. for downloads see `http://www.fz-juelich.de/nic-series/volume10/doltsinis.pdf`.
28. N. L. Doltsinis and D. Marx. *J. Theor. Comp. Chem.*, 1:319–349, 2002.
29. M. Sprik. *Chem. Phys.*, 258:139, 2000.
30. J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, and M. Sprik. *J. Am. Chem. Soc.*, 124:6594, 2002.
31. N. L. Doltsinis and M. Sprik. *Phys. Chem. Chem. Phys. Lett.*, 5:2612, 2003.
32. H. Langer and N. L. Doltsinis. *J. Chem. Phys.*, 118:5400, 2003.
33. H. Langer. Diplomarbeit, Ruhr-Universität Bochum, 2002.